

IMPERIAL COLLEGE LONDON

BSc FINAL PROJECT

---

# Galaxy Clustering Effects in Weak Gravitational Lensing Statistics

Fourier Space Analysis

---

*Author:*

Sofia Gonzalez Garcia  
CID: 01490427

*Supervisor:*

Prof. Alan Heavens

*Assessor:*

Prof. Andrew Jaffe

*Word Count:*  
5300

*A thesis submitted in fulfillment of the requirements  
for the degree of Physics with Theoretical Physics  
in the*

Blackett Laboratory, Department of Physics

January 2021



## DECLARATION AND ACKNOWLEDGEMENTS

---

This project has been developed during the first term of the academic year 2020-2021. This report aims to summarise my work and findings, together with suggestions and considerations for future research.

This work has been a shared experience with my project partner, Amol Bhuptani. All of the computational methods have been attempted independently, each of us using our own individual approaches. However, we have worked in a cooperative fashion, discussing results and comparing computational procedures. I have mostly focused on the empirical computation of the power spectra of convergence and shear; while my partner has also explored the implications of the choice of protocol for empty bin filling.

All graphs in this report have been obtained by myself unless stated otherwise.

I would like to express my thankfulness to Prof. Alan Heavens for his guidance and support during this project, and to Amol for their invaluable suggestions, help and patience.

*Sofia Gonzalez Garcia*

## ABSTRACT

---

The clustering of sources whose shape is measured in weak gravitational lensing surveys is a systematic effect which could impact the estimated values of cosmological parameters. Previous weak lensing analysis had ignored the implications of source clustering because its predicted impact was negligible compared to the error bars of the surveys where they were used. This project signifies the first attempt to quantify the effect of source clustering.

For this work, we created our own weak lensing computational simulation in a framework of strong approximations and assumptions. The simulation obtained the power spectra of convergence and shear for different sets of galaxies. These were then compared to the true spectra, which was yielded from a set of gridded sources with no intrinsic shape error. The method for obtaining and comparing these power spectra was primarily numerical and empirical as suggested by our supervisor. However, it was supported by analytical derivations of Fourier space parameters, which provided an alternative method for understanding source clustering effects and a deeper insight. The data obtained for the different sets of sources were then statistically compared using the chi squared distribution,  $\chi^2$ . We found that the power spectra of clustered sources does not deviate from those of the true field when the variance of the shape noise (or shot noise) is considered. However, clustering generates larger uncertainties and a lower signal-to-noise ratio. These conclusions are only valid when the empty bins have been filled by adding one source.

We conclude that more thorough analysis needs to be done to understand how clustering affects the estimation of the cosmological parameters that will be inferred in the upcoming surveys. Furthermore, we suggest further implementations that could be considered in future research, such as revisiting certain approximations we have made in this project.

## CONTENTS

1	INTRODUCTION	1
2	THEORETICAL BACKGROUND	3
2.1	Cosmological model . . . . .	3
2.2	Light deflection and lensing equations . . . . .	3
2.3	Convergence and Shear . . . . .	5
2.4	Structure formation . . . . .	5
2.4.1	Galaxy bias . . . . .	6
3	STATISTICAL TREATMENT	7
3.1	Two-point correlation function . . . . .	7
3.2	Power spectrum . . . . .	7
3.3	Comparison method: chi squared . . . . .	8
3.4	Parameter propagation in Fourier transforms . . . . .	8
3.4.1	Transforming error, $n_s^b$ : . . . . .	9
3.4.2	Transforming the mean, $\mu_s$ : . . . . .	9
4	METHOD	11
4.1	Gravitational potential creation . . . . .	14
4.2	Source creation . . . . .	15
4.2.1	Poisson sampling procedure . . . . .	15
4.3	Integration method: obtaining $\gamma$ and $\kappa$ . . . . .	16
4.4	Real Space analysis: Correlation function . . . . .	17
4.5	Fourier Space analysis: empirical approach . . . . .	17
4.6	Fourier Space analysis: analytical approach . . . . .	22
5	RESULTS	24
5.1	Empirical power spectra results . . . . .	25
5.2	Fourier amplitude comparison results . . . . .	28
5.3	Signal-to-noise ratio . . . . .	29
6	CONCLUSIONS	31
	BIBLIOGRAPHY	33

## INTRODUCTION

---

Gravitational lensing is the distortion of images of distant galaxies due to the inhomogeneous gravitational field in which the light rays from these galaxies propagate. The inhomogeneity of the universe arises from the density fluctuations of its large-scale structure (LSS) [1]. The amount of bending of these light rays due to the presence of intervening matter is described by Einstein's theory of General Relativity [3].

In the study of gravitational lensing we can consider two regimes depending on the magnitude of the distortion: strong lensing and weak lensing. The former corresponds to strongly non-linear perturbations caused by highly non-linear massive objects [4]. On the other hand, in the weak lensing regime (also called cosmic shearing), the order of the distortion is of a few percent, much smaller than the width of the intrinsic shape and size distribution of the sources [1]. Because the effects are so subtle, weak lensing requires a purely statistical treatment by using galaxy shear and magnification correlations and, therefore, the use of a large set of sources. Weak gravitational lensing will be the regime of focus in our project.

The main observational advantage of weak lensing is that it directly probes the distribution of matter in the Universe [4]. Cosmic shearing is widely used as a probe for the LSS, ranging from the highly non-linear, non-Gaussian sub-megaparsec (Mpc) regime, out to very large, linear scales of more than a hundred Mpc. It can be used, for example, to shed light on the effect of dark energy on the growth of structure [1], or to constrain theories of modified gravity and test Einstein's General Relativity on cosmological scales [5]. Weak lensing analysis is also commonly used alongside astronomical surveys, yielding valuable information about galaxy formation processes. Surveys to date have already been able to impose strong constraints on  $\Omega_M$  (the matter density relative to critical),  $\sigma_8$  (the amplitude of mass fluctuations) and on the equation of state of dark energy [6].

Previous astronomical surveys had ignored the clustering effect of the sources whose distortion was studied. However, astronomical surveys employing weak lensing techniques have undergone major improvements in the recent years, with a significant increase in precision. Ongoing and planned surveys, such as Euclid, the Large Synoptic Survey Telescope (LSST) or the Supernova/Acceleration Probe (SNAP) [2] will achieve small statistical uncertainties because of the larger coverage of sky and increment in number of galaxies, potentially increasing the amount of data by a factor of 100 [1]. Therefore, the current challenge is to reduce systematic errors, and so the discussion regarding the significance of

clustering effects needs to be reopened.

In this project we studied the effects of galaxy clustering on the power spectra of convergence and shear, comparing them to those of gridded sources. The method used for this investigation is fully computational. The statistical Fourier analysis suggested by our supervisor was supported by an alternative Fourier parameter propagation analysis.

## THEORETICAL BACKGROUND

---

This sections provides an overview of the main theoretical aspects behind weak gravitational lensing. Thorough derivations can be found in Chapter 10 of *Modern Cosmology* by Dodelson or in *Weak Gravitational Lensing* by Bartelmann and Schneider.

### 2.1 COSMOLOGICAL MODEL

Viewed on sufficiently large scales, the spatial distribution of matter in the universe is homogeneous and isotropic, as suggested by the cosmological principle [7]. This is clearly not valid at small scales, where matter is distributed in clumps separated by large voids, but it holds beyond the homogeneity scale, estimated to be  $200 h^{-1} \text{Mpc}$  [7]. The metric that satisfies the cosmological principle, including first order perturbations, is of the form [3]:

$$ds^2 = \left(1 + \frac{2\Psi}{c^2}\right) c^2 dt^2 - a^2(t) \left(1 - \frac{2\Phi}{c^2}\right) dl^2, \quad (1)$$

where  $dl$  is the line element of the homogeneous and isotropic three-space,  $a(t)$  is a scale factor (set to unity at present time  $t = t_0$ ) and  $\Psi$ ,  $\Phi$  are two scalar metric perturbations describing weak fields such that  $\Psi, \Phi \ll c^2$ . The most general form for the spatial line element is:

$$dl^2 = d\chi^2 + f_K^2(\chi) d\omega^2, \quad (2)$$

where  $\chi$  is the comoving coordinate,  $f_K^2$  is the the comoving angular distance and  $d\omega$  is the angular element. Considering homogeneity constraints, the allowed forms for  $f_K^2$  are:

$$f_K^2(\chi) = \begin{cases} K^{-\frac{1}{2}} \sin(K^{\frac{1}{2}} \chi) & \text{if } K > 0 \text{ (spherical)} \\ \chi & \text{if } K = 0 \text{ (flat)} \\ -K^{-\frac{1}{2}} \sinh(-K^{\frac{1}{2}} \chi) & \text{if } K < 0 \text{ (hyperbolic)} \end{cases} \quad (3)$$

### 2.2 LIGHT DEFLECTION AND LENSING EQUATIONS

When propagating in an inhomogeneous gravitational potential, light rays are deflected [1]. Consider the trajectory of a light rays deflected by a gravitational potential  $\Phi$  as depicted in Figure 1.



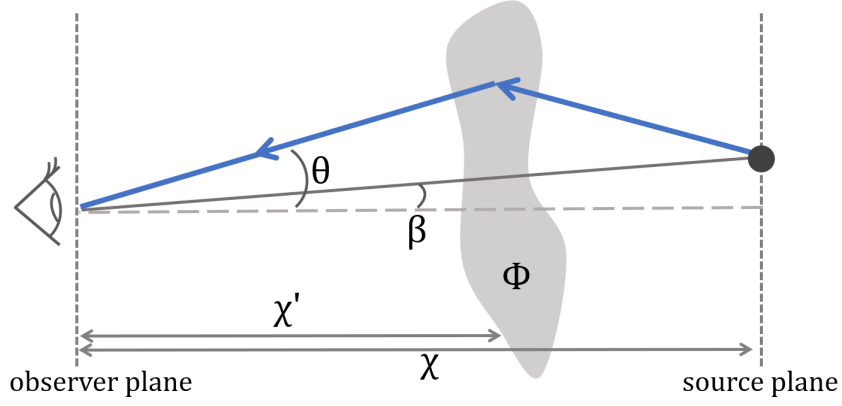


Figure 1: Diagram representing the deflection of a source by a potential fluctuation  $\Phi$  at a distance  $\chi'$  from the observer. The gray solid line represents the unperturbed trajectory, subtended at an angle  $\beta$  from the horizontal. The blue line shows the deflected path of the light ray, converging on the observer under the angle  $\theta$ .

We can define the scaled deflection angle,  $\vec{\alpha} = \vec{\theta} - \vec{\beta}$ . The expression for  $\vec{\alpha}$  can be obtained from GR considerations and application of Fermat's principle, yielding [4]:

$$\vec{\alpha} = \frac{2}{c^2} \int_0^\chi d\chi' \frac{f_k(\chi - \chi')}{f_k(\chi)} \nabla_\perp \Phi(\chi'). \quad (4)$$

Here,  $\nabla_\perp \Phi$  represents the gradient of the potential in the direction perpendicular to the deflected trajectory. In the weak lensing regime, this deflection angle will be assumed to be small and we can make use of the Born approximation [4], by which we approximate the gravitational potential along the deflected trajectory to be the potential along the undeflected trajectory. This implies that a change in angle  $\vec{\theta}$  corresponds to a transverse distance  $f_k(\chi') \vec{\theta}$ . It is also useful to define the lensing potential,  $\psi$ , such that the scaled deflection angle is just the gradient of the potential [1]:

$$\nabla \psi = \vec{\alpha} \rightarrow \psi(\theta, \chi) = \frac{2}{c^2} \int_0^\chi d\chi' \frac{f_k(\chi - \chi')}{f_k(\chi) f_k(\chi')} \Phi(\chi', f_k(\chi') \vec{\theta}). \quad (5)$$

We can define the amplification matrix as the Jacobian between the unlensed and lensed coordinate systems [1]:

$$\vec{\beta} = \mathcal{A} \vec{\theta} \rightarrow \mathcal{A} = \frac{\partial \vec{\beta}}{\partial \vec{\theta}} = \delta - \frac{\partial \vec{\alpha}}{\partial \vec{\theta}}, \quad (6)$$

where  $\delta$  is the Kronecker delta and

$$\frac{\partial \vec{\alpha}}{\partial \vec{\theta}} = \frac{2}{c^2} \int_0^\chi d\chi' \frac{f_k(\chi - \chi') f_k(\chi')}{f_k(\chi)} \partial_i \partial_j \Phi(\chi', f_k(\chi') \vec{\theta}), \quad (7)$$

with the derivatives being taken along the unperturbed trajectory.

## 2.3 CONVERGENCE AND SHEAR

The symmetrical matrix  $\mathcal{A}$  can be parametrized in terms of the scalar convergence,  $\kappa$ , and the two-component spin-two shear,  $\gamma = \gamma_1 + i\gamma_2$  [1]:

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa & 0 \\ 0 & 1 - \kappa \end{pmatrix} + \begin{pmatrix} -\gamma_1 & -\gamma_2 \\ -\gamma_2 & \gamma_1 \end{pmatrix}. \quad (8)$$

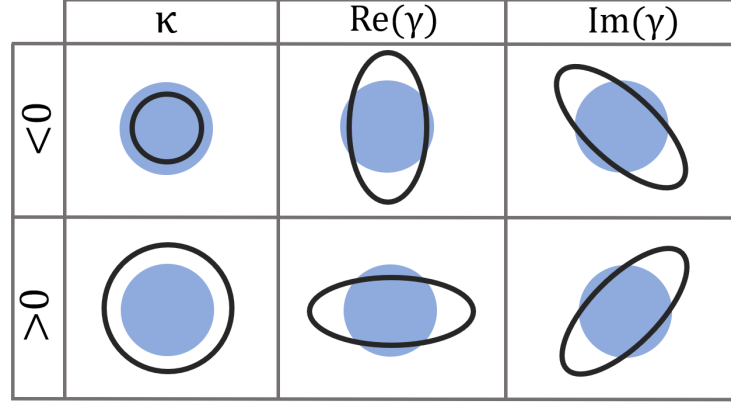


Figure 2: Diagram representing in black the effect of the components of convergence and shear on a circular source represented by the solid blue circle. Convergence indicates magnification and shear describes the area-preserving distortion [3].

The convergence and shear can therefore be rewritten in terms of the lensing potential  $\psi$ :

$$\text{convergence: } \kappa = \frac{1}{2}(\partial_1^2 + \partial_2^2)\psi, \quad (9)$$

$$\text{shear: } \gamma = \gamma_1 + i\gamma_2 = \frac{1}{2}(\partial_1^2 - \partial_2^2 + 2i\partial_1\partial_2)\psi, \quad (10)$$

where the derivatives are with respects to  $\vec{\theta}$ .

## 2.4 STRUCTURE FORMATION

Fluctuations of the matter density,  $\rho$ , around the mean density,  $\bar{\rho}$ , are parametrised by the density contrast  $\delta$  (or overdensity), given by [1]:

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}. \quad (11)$$

The gravitational potential sourced by density fluctuations,  $\Phi(\mathbf{r})$ , is related to the density contrast via the Poisson equation [1]:

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho} \delta, \quad (12)$$

where  $a$  is the scale factor as described in section 2.1. We can relate  $\delta$  to its power spectrum [9] through the expression:

$$P_k = \langle |\delta_k|^2 \rangle. \quad (13)$$

The shape of the power spectrum will vary significantly depending on the damping of fluctuations or the type of dark matter [7]. Figure 3 depicts the linear matter power spectrum inferred from different cosmological probes.

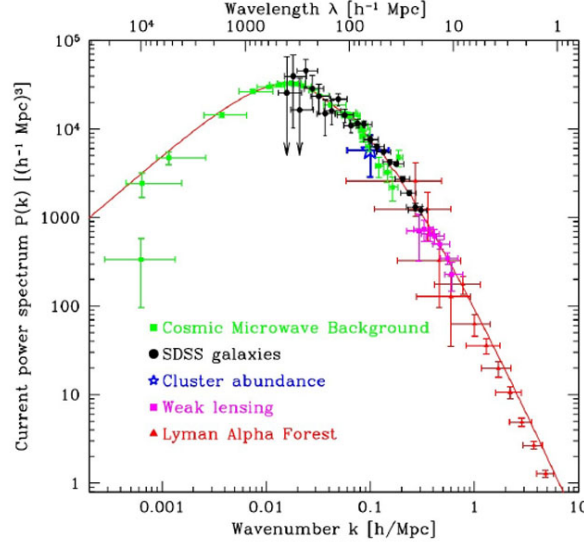


Figure 3: (Taken from [8]) The figure shows the linear matter power spectrum  $P(k)$  versus wavenumber from various measurements of cosmological structure. The best fit  $\Lambda$ CDM model is shown as a solid line [9].

#### 2.4.1 Galaxy bias

There is no reason to assume that the distribution of light in the universe traces exactly the distribution of mass. On large scales where density fluctuations are small, a commonly used model is to assume the existence of a linear bias  $b$  between fluctuations in mass and in the galaxy distribution [11], such that:

$$\left(\frac{\delta\rho}{\rho}\right)_{\text{galaxies}} = b \left(\frac{\delta\rho}{\rho}\right)_{\text{mass}}, \quad (14)$$

where  $b$  is the galaxy bias factor <sup>1</sup>.

<sup>1</sup> Later works (Hamilton et al., 2000) have found that this bias is scale-dependent [11].

## STATISTICAL TREATMENT

---

As discussed previously, the effects of convergence and shear ( $\kappa, \gamma \approx 0.01$ ) on the magnification and ellipticities of galaxies is orders of magnitude smaller than the dispersion of the unlensed intrinsic shape of the galaxies,  $\epsilon_s$ , which has a value of around  $\sigma_s \sim 0.4$  [4]. The value for magnification and ellipticity that we measure (the signal) therefore are [1]:

$$\text{for shear: } \epsilon_\gamma \simeq \epsilon_{s1} + \gamma, \quad (15)$$

$$\text{for convergence}^1: \epsilon_\kappa \simeq \epsilon_{s2} + \kappa, \quad (16)$$

where  $\epsilon_{s1/s2} \sim \mathcal{N}(0, \sigma_{s1/s2}^2)$ . Evidently, a large number of galaxies  $N$  is required to decrease the noise, which scales as  $\sim \frac{\epsilon_{s1/s2}}{\sqrt{N}}$  [4].

### 3.1 TWO-POINT CORRELATION FUNCTION

The two-point correlation function measures the relationship between the magnification or ellipticities of galaxy pairs separated by a fixed distance. Its expression is given by [4]:

$$\xi_\gamma(\vec{r}) = \langle \epsilon_\gamma(\vec{x}) \epsilon_\gamma(\vec{x} + \vec{r}) \rangle_{\vec{x}}, \quad (17)$$

$$\xi_\kappa(\vec{r}) = \langle \epsilon_\kappa(\vec{x}) \epsilon_\kappa(\vec{x} + \vec{r}) \rangle_{\vec{x}}. \quad (18)$$

However, when comparing the correlation function for two sets of galaxies, we cannot conclude they are statistically different since points in this statistic are correlated [13]. Therefore, further treatment would be needed, and we will opt to focus on the power spectrum analysis instead.

### 3.2 POWER SPECTRUM

Expressed in Fourier space, the two-point correlation functions defines the power spectrum  $P_{\gamma/\kappa}(k)$  which reads [1]:

$$\langle \tilde{\gamma}(\vec{k}) \tilde{\gamma}^*(\vec{k}') \rangle = (2\pi)^2 P_\gamma(k) \delta_D(\vec{k} - \vec{k}'), \quad (19)$$

$$\langle \tilde{\kappa}(\vec{k}) \tilde{\kappa}^*(\vec{k}') \rangle = (2\pi)^2 P_\kappa(k) \delta_D(\vec{k} - \vec{k}'), \quad (20)$$

where  $\tilde{\gamma}/\tilde{\kappa}$  and  $\gamma^*/\kappa^*$  express the variable in phase space and complex conjugation respectively, and  $\delta_D$  is the Dirac-delta function.

---

<sup>1</sup> We do not have a good model for convergence noise, so we assume it works in a similar fashion to shear.

### 3.3 COMPARISON METHOD: CHI SQUARED

To study the statistical difference between the expected theoretical data,  $\bar{d}_i$  and observed data,  $d_i$ , we will use the chi squared distribution. If all the data points are independent <sup>2</sup>, the chi squared is defined as [12]:

$$\chi^2 = \sum_i \frac{(d_i - \bar{d}_i)^2}{\sigma_i^2}, \quad (21)$$

where the  $\sigma_i$  are the uncertainties on  $d_i$  and  $\bar{d}_i$ .

For a good fit between the theoretical and observed values, we expect  $d_i$  and  $\bar{d}_i$  to differ by amounts similar to  $\sigma_i$ , since they are expected to follow a Gaussian distribution. Therefore, the value of  $\chi^2$  is expected to be  $\chi^2 \simeq N_{\text{dof}} \pm \sqrt{2N_{\text{dof}}}$ , where  $N_{\text{dof}}$  are the number of degrees of freedom, defined as  $N_{\text{data}} - N_{\text{parameters}}$ .

### 3.4 PARAMETER PROPAGATION IN FOURIER TRANSFORMS

This section will introduce the theoretical background we suggested in order to treat certain aspects of the investigation analytically. It was used alongside our empirical investigations to offer more insight.

Consider a set of galaxies distributed in a square in real space <sup>3</sup>, where each of the galaxies has individual values for shear and convergence. We proceed to bin these galaxies in equally sized boxes.

Take one of these boxes, which contains  $g_b$  galaxies. The value of the shear or convergence,  $\epsilon$ , of one of the sources,  $s$ , in the bin,  $b$ , will be:

$$\epsilon_s^b = \mu_s^b + n_s^b \quad (22)$$

where  $s \in \{1, 2, \dots, g_b\}$ ,  $n = \mathcal{N}(0, \sigma_b)$  and  $\mu_s^b$  is the true value of the converge and shear at the grid point ( $\gamma$  or  $\kappa$ ) without added intrinsic shape noise. It is clear that as the number of galaxies in a bin increases, the variance will be reduced, and so:

$$\sigma_b^2 = \frac{\sigma_s^2}{g_b} \quad (23)$$

where  $\sigma_s$  is the standard deviation for the distribution of intrinsic ellipticities of the galaxies, as used in Equations 15 and 16.

We will now see how the parameters in Equation 22 transform to Fourier space.

The results stated below <sup>4</sup> are subject to the following assumptions:

- 
- <sup>2</sup> If the observed data points were correlated, we would need to consider the off-diagonal terms of the error matrix [12].
  - <sup>3</sup> The motivation for this procedure and the details of it will be explained in the METHOD section.
  - <sup>4</sup> These derivations are based on Dodelson, *Gravitational Lensing* Chapter 8 Section 5 [13].

1. There is at least one source in each bin.
2. The Fourier modes are independent of each other, i.e uncorrelated across scales. We ignore the off diagonal terms in the noise matrix for the amplitudes.  
If the amplitudes were correlated, the derivations would have to be adapted to include covariance [1].
3. The mass distribution used in simulation is given by a random Gaussian field.

#### 3.4.1 Transforming error, $n_s^b$ :

When transforming  $e_s^b$  to Fourier space, it can be shown that [4, 13]:

$$\text{Var}[\mathcal{F}(e_s^b)] = \text{Var}[\tilde{e}_{\vec{k}}] = \tilde{\sigma}_{\vec{k}}^2 = \sum_b \sigma_b^2, \quad (24)$$

where the sum is over all bins in real space and clearly  $\tilde{e}_{\vec{k}}$  are the Fourier amplitudes of the convergence and shear signal in Fourier space. This result shows that the shape noise,  $\tilde{\sigma}_{\vec{k}}^2$ , in Fourier space is independent of  $\vec{k}$ . It is indeed the same for all modes.

This means that the error in Fourier space,  $\tilde{n}_{\vec{k}}$  will be drawn from a normal distribution  $\mathcal{N} \sim (0, \tilde{\sigma}_{\vec{k}}^2)$ . Incidentally, for the same amount of total galaxies, the configuration that will minimise  $\tilde{\sigma}_{\vec{k}}^2$  is that in which the galaxies are equally distributed among the boxes<sup>5</sup>.

#### 3.4.2 Transforming the mean, $\mu_s$ :

Due to our random Gaussian field assumption, the distribution of the mean in Fourier space,  $\tilde{\mu}_{\vec{k}}$ , is given by [13]:

$$\tilde{\mu}_{\vec{k}} \sim \mathcal{N}(0, \tilde{\sigma}_{\mu}^2), \quad (25)$$

where  $\tilde{\sigma}_{\mu}$  has the same value for all the  $\vec{k}$  with equal  $|\vec{k}|$ , meaning:

$$\tilde{\sigma}_{\mu}^2 = \tilde{\sigma}_{\mu}^2(k). \quad (26)$$

We could obtain the value of  $\tilde{\sigma}_{\mu}$  analytically but this would require us to study the mass power spectrum in more detail in order to understand how it generates the convergence and shear fields. Due to time limitations we did not consider this procedure in our analysis.

Summarising our results, when transforming  $e_s^b$  to Fourier space, we get:

$$\mathcal{F}[e_s^b] = \mathcal{F}[\mu_s^b] + \mathcal{F}[n_s^b] = \mathcal{N}(0, \tilde{\sigma}_{\mu}^2) + \mathcal{N}(0, \tilde{\sigma}_{\vec{k}}^2). \quad (27)$$

<sup>5</sup> This can be easily proved using the method for Lagrange multipliers. This configuration corresponds to the configuration with the largest statistical weight.

We can define the estimator:

$$\hat{\sigma}_T^2 = \hat{\sigma}_\mu^2 + \hat{\sigma}_k^2, \quad (28)$$

where  $\hat{\sigma}_T^2$  is just a function of  $|\vec{k}|$ . Because we know the value of  $\tilde{\sigma}_k$ , we simply set  $\hat{\sigma}_k = \tilde{\sigma}_k$ .

Because the total variance,  $\hat{\sigma}_T^2$ , only depends on  $|\vec{k}|$  [13], it can be estimated by the average signal in an annuli of k-space, which contains all the Fourier modes (indexed with n) whose  $|\vec{k}|$  ranges from  $|\vec{k}| \rightarrow |\vec{k}| + \delta k$ . The expression for the estimator of the total variance is:

$$\hat{\sigma}_T^2 = \frac{1}{g_k} \sum_n |\tilde{\epsilon}_k(\vec{k}_n)|^2, \quad (29)$$

where  $g_k$  is the number of Fourier modes in the annulus. The total variance  $\tilde{\sigma}_T^2$  takes the same form as the power spectrum.

Therefore, it can be expressed as:

$$\hat{\sigma}_T^2 \sim \mathcal{N}(\tilde{\sigma}_T^2, \text{Var}[\tilde{\sigma}_T^2]) \quad (30)$$

where we rewrite the variance to be  $\text{Var}[\tilde{\sigma}_T^2] =$

$$\text{Var} \left[ \frac{\tilde{\sigma}_T^2}{g_k} \sum_n \frac{|\tilde{\epsilon}_k(\vec{k}_n)|^2}{\tilde{\sigma}_T^2} \right] = \frac{\tilde{\sigma}_T^4}{g_k^2} \text{Var} [\chi^2(g_k)] = \frac{1}{g_k} \tilde{\sigma}_T^4, \quad (31)$$

where  $\chi^2$  is the chi squared distribution.

The final distribution for  $\hat{\sigma}_T^2$  is:

$$\hat{\sigma}_T^2 \sim \mathcal{N} \left( \tilde{\sigma}_T^2, \frac{1}{g_k} \tilde{\sigma}_T^4 \right). \quad (32)$$

Rearranging  $\hat{\sigma}_\mu^2$  in Equation 28:

$$\hat{\sigma}_\mu^2 = \hat{\sigma}_T^2 - \tilde{\sigma}_k^2 \sim \mathcal{N} \left( \tilde{\sigma}_\mu^2, \frac{1}{g_k} (\hat{\sigma}_\mu^2 + \tilde{\sigma}_k^2)^2 \right), \quad (33)$$

which is an expression that we can use to determine the ratio of power spectrum signal,  $\hat{\sigma}_\mu^4$ , to noise  $\text{Var}[\hat{\sigma}_\mu^2]$ :

$$\frac{\hat{\sigma}_\mu^4}{\text{Var}(\hat{\sigma}_\mu^2)} = \frac{g_k \hat{\sigma}_\mu^4}{(\hat{\sigma}_\mu^2 + \tilde{\sigma}_k^2)^2} = \frac{g_k (\hat{\sigma}_T^2 - \tilde{\sigma}_k^2)^2}{\hat{\sigma}_T^4}. \quad (34)$$

## METHOD

This chapter will describe the method and procedures behind our investigation. This project was mainly computational and involved the creation of weak gravitational lensing routines using Python. Although we mostly focused on the procedures suggested by our supervisor, we also carried out a separate analytical Fourier space investigation, whose motivation and procedure will be explained in Section 4.6.

For the computational procedures to follow, the subsequent approximations were made:

1. Flat universe:  $f_K(\chi) = \chi$  (see Equation 3).
2. Born approximation, as described by the theory. We will therefore integrate along the unperturbed light ray trajectory.
3. We ignore gradients along the  $z$ -direction because we consider a long sided rectangular box of the Universe, as seen below.

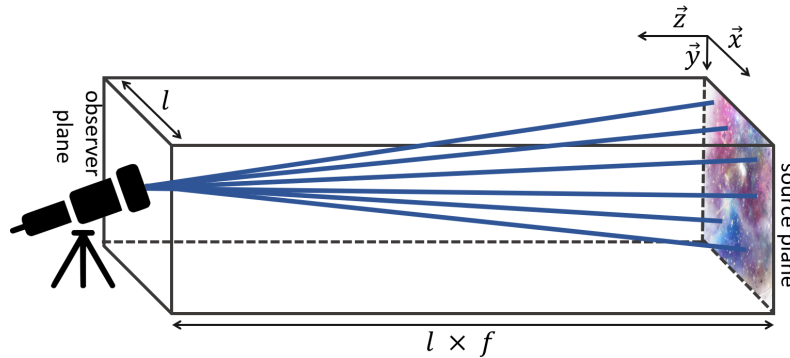


Figure 4: Schematic representation of the rectangular section of the Universe used in our simulation. The factor  $f$  represents how much larger the long side is with respects to the square side. In our simulation  $f = 5$ , this way we can use small angle approximations.

4. Neglect any time evolution or redshift dependence of light deflection. Ignoring this time dependence will produce a less realistic simulation. Literature suggests a maximum of 300Mpc should be used for a given patch of sky, and that larger simulations should be split into smaller boxes [1].
5. All of our sources are on the same plane at the same red-shift.
6. We used a random Gaussian field and we probed in the linear domain, since we used the power spectrum in Figure 13. Due to



the Gaussian nature of the field, all the information is contained in the correlation function (real space) and power spectrum (Fourier space), so we need not consider other parameters.

7. No correlation between galaxies intrinsic alignments, i.e.  $\langle \epsilon_s(\mathbf{x}) \epsilon_s(\mathbf{x} + \mathbf{r}) \rangle = 0$  and  $\langle \epsilon_s(\mathbf{x}) \gamma / \kappa(\mathbf{x} + \mathbf{r}) \rangle = 0$ . This reduces Equations 17 and 18 to:

$$\xi_\gamma(\mathbf{r}) = \langle \epsilon_\gamma(\mathbf{x}) \epsilon_\gamma(\mathbf{x} + \mathbf{r}) \rangle_{\mathbf{x}} = \langle \gamma(\mathbf{x}) \gamma(\mathbf{x} + \mathbf{r}) \rangle_{\mathbf{x}} \quad (35)$$

$$\xi_\kappa(\mathbf{r}) = \langle \epsilon_\kappa(\mathbf{x}) \epsilon_\kappa(\mathbf{x} + \mathbf{r}) \rangle_{\mathbf{x}} = \langle \kappa(\mathbf{x}) \kappa(\mathbf{x} + \mathbf{r}) \rangle_{\mathbf{x}} \quad (36)$$

8. We neglected galaxy-galaxy lensing effects, i.e. no lensing between sources.
9. We considered the dispersion of the error associated with the intrinsic shape to be equal for convergence and shear, i.e.  $\sigma_{s1} = \sigma_{s2} = \sigma_s$  in Equations 15 and 16. As long as this value is consistent across galaxy sets, our comparison will be unchanged.  
Note that because shear is composed of a real and imaginary part, the standard deviation will split as:  $\sigma_{r/i} = \frac{\sigma_s}{\sqrt{2}}$ .
10. We ignore errors that might be relevant in empirical analyses, such as those encoded in the *point spread function* [13] or masking due to objects in the light ray path.

Using approximations 1,2 and 3, our expressions for  $\kappa$  and  $\gamma$  (Equations 9 and 10) in terms of the lensing potential  $\psi$  simplify to:

$$\kappa = \frac{1}{c^2} (\partial_x^2 + \partial_y^2) \int_0^r \frac{(\mathbf{r} - \mathbf{r}')}{r} r' \Phi(\mathbf{r}') d\mathbf{r}' \quad (37)$$

$$\gamma = \frac{1}{c^2} (\partial_x^2 - \partial_y^2 + 2i \partial_x \partial_y) \int_0^r \frac{(\mathbf{r} - \mathbf{r}')}{r} r' \Phi(\mathbf{r}') d\mathbf{r}' \quad (38)$$

In our algorithms, for convenience, we will take derivatives in Fourier space. The Fourier form of Equations 37 and 38 is:

$$\kappa = \frac{1}{c^2} \int_0^r d\mathbf{r}' f(\mathbf{r}') \mathcal{F}^{-1} [(-k_x^2 - k_y^2) \mathcal{F}[\Phi(\mathbf{r})]] \quad (39)$$

$$\gamma = \frac{1}{c^2} \int_0^r d\mathbf{r}' f(\mathbf{r}') \mathcal{F}^{-1} [(-k_x^2 + k_y^2 - 2ik_x k_y) \mathcal{F}[\Phi(\mathbf{r})]] \quad (40)$$

where  $f(\mathbf{r}') = r' \frac{\mathbf{r} - \mathbf{r}'}{r}$ .

For the reader to follow the computational procedures more conveniently, a flow chart of the method is summarised in Figure 5:

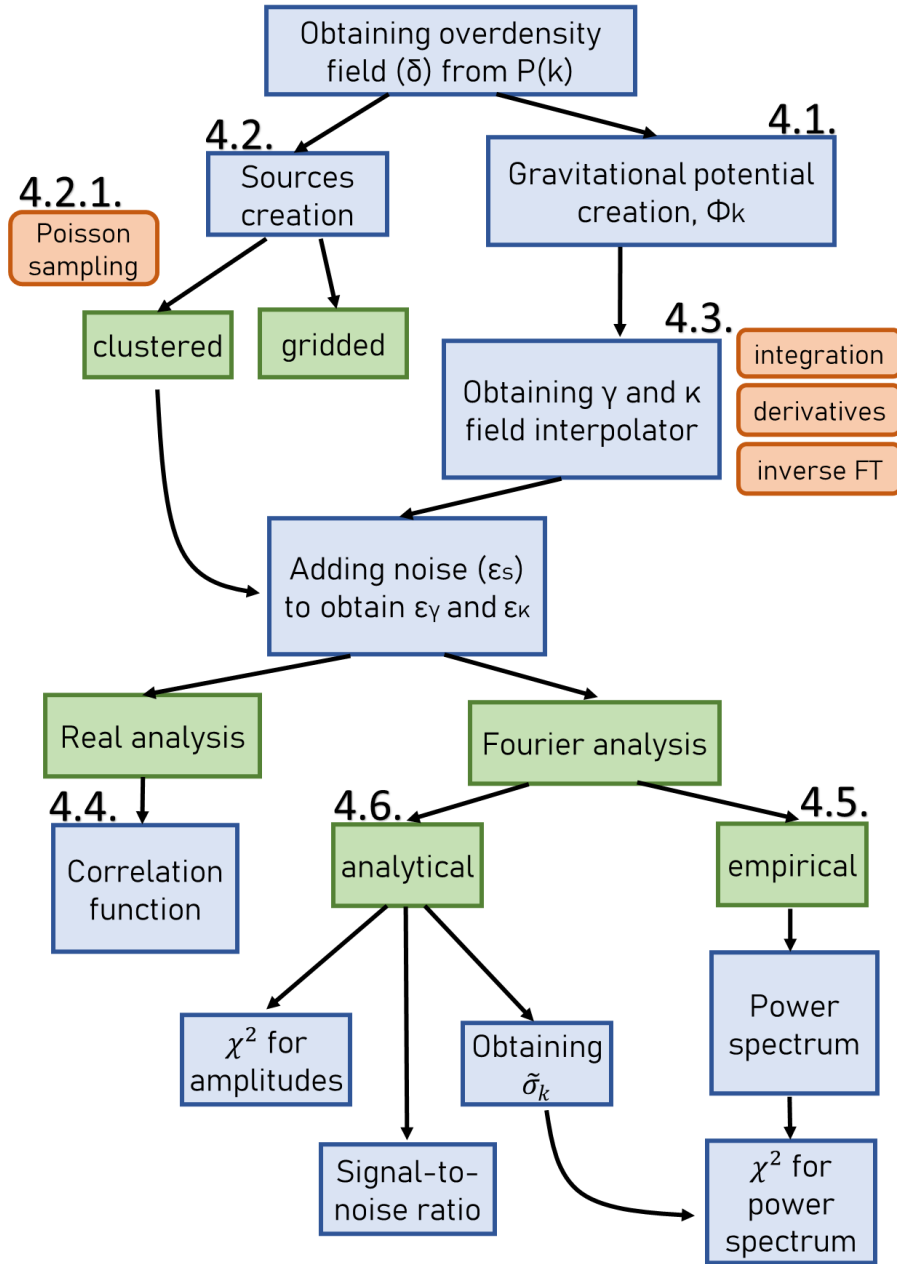


Figure 5: Flow chart of the computational procedures that will be described below. The number represents the section's identification. Blue: processes, Orange: functions used, Green: rest.

The success of the computational procedures was determined by comparing the obtained graphs with those of my partner and through sev-

eral checks that will be described below in the procedure's respective section.

Consulting the documentation for the prewritten Python methods, such as *numpy*'s Fourier transform library, was crucial in order to take care of the conventions and factors to be considered.

#### 4.1 GRAVITATIONAL POTENTIAL CREATION

Our objective with this routine was to obtain a random Gaussian field based on the power spectrum presented in Figure 3. This field acted as our gravitational potential for the rest of the simulation.

We aimed to generate a large *box* of the Universe (larger than 200x200 Mpc<sup>2</sup> on the square side) without compromising on the resolution of the simulation (i.e. the number of data points). Because Python built-in Discrete Fourier Transform (DFT) methods work in powers of 2, we chose our lengths to be powers of 2 for added efficiency.

We created 3 arrays for each of the 3 directions in k-space:  $k_x, k_y, k_z$ . The spacing along each  $\vec{k}$  direction was  $\frac{2\pi}{L}$ , where L is the length of each side of the box in real space [10]. We then generated a 3D matrix with the values of the magnitude of the k vector,  $|k|$ , at every point in k-space. We obtained the matter power spectrum value,  $P_k$  corresponding to each  $|k|$  using the power spectrum in Figure 3. The overdensity field in k-space is then calculated using Equation 13, giving:

$$\delta(\mathbf{k}) = \delta_r + i\delta_i = \mathcal{N}\left(0, \frac{P_k}{2}\right) + i\mathcal{N}\left(0, \frac{P_k}{2}\right). \quad (41)$$

We set the  $k = 0$  amplitude to be 0, in order for the final field to have a mean of zero. Using the Poisson Equation (see Equation 12) in k-space we finally produced the gravitational potential  $\Phi(\mathbf{k})$ .

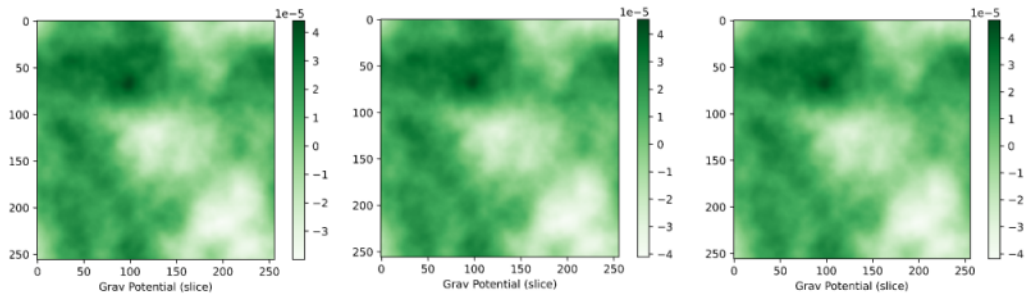


Figure 6: Three consecutive slices (x-y planes) of the 3D gravitational potential we created, which is clearly continuous. In our simulation, the long side of the potential box is a factor of 5 times as large as the sides of the square plane. Distances are in Mpc.

## 4.2 SOURCE CREATION

For our simulation, we need to create 3 different types of sources:

1. **Gridded galaxies**, which we take as the *true* background sources, and their correlation function and power spectrum was be considered to be the theoretical expectation.  
Gridded galaxies can be understood as the limit of taking an infinite number of random galaxies.
2. **Random galaxies**, whose positions are completely uncorrelated.
3. **Clustered galaxies**, whose distribution was drawn from the matter power spectrum of the Universe (Figure 3) using a technique called Poisson sampling (see below). We exploited the idea of *galaxy bias* factor (discussed in section 2.4.1) to tune the level of clustering of our sources and compare the effect of this parameter on our results.

### 4.2.1 Poisson sampling procedure

For every point in a fine grid on the source plane we draw a number from a Poisson distribution, which corresponded to the number of galaxies at that position. We set the mean of the Poisson distribution to be:

$$\mu_P = F(1 + \delta(\mathbf{r})) \propto \rho(\mathbf{r}) \quad (42)$$

where  $F$  is a factor of our choice which controls the galaxy density and  $\delta(\mathbf{r})$  is the overdensity field in position space. The probability of getting more than one galaxy in the same position was negligible,  $\mu_P \ll 1$ . A check that this routine was working as expected can be found below in Figure 7.

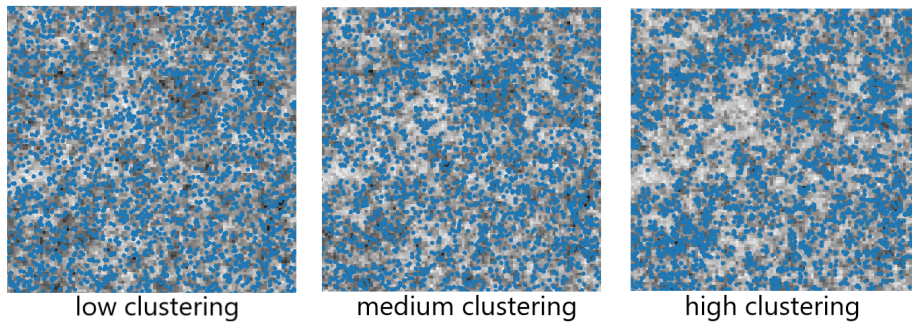


Figure 7: Gray: mass overdensity field. Blue: galaxy positions.

Detail of different levels of galaxy clustering obtained from the same mass power spectrum by tuning its range. This test suggests that our clustered sources creation routine works as expected.

4.3 INTEGRATION METHOD: OBTAINING  $\gamma$  AND  $\kappa$ 

The integration was carried out in Fourier space for convenience, using Equations 37 and 38 in their Fourier form. Because we were taking the Born approximation, we simply integrated along the unperturbed trajectory for each light ray. This was worked out by calculating the *boxes* of gravitational potential that the light ray will cross on their trajectory to the observer, and adding all the contributions of the integrand.

We can tune the resolution of our integration (the precision with which we choose the crossed boxes) in order to increase the speed of computation.

The integration was only be performed for the gridded galaxies. We then created an interpolator function which we used to obtain the convergence and shear values for each source in the random and clustered galaxy sets. This increased the computational efficiency significantly. Plots of the convergence and shear fields obtained with this routine are depicted in Figure 8.

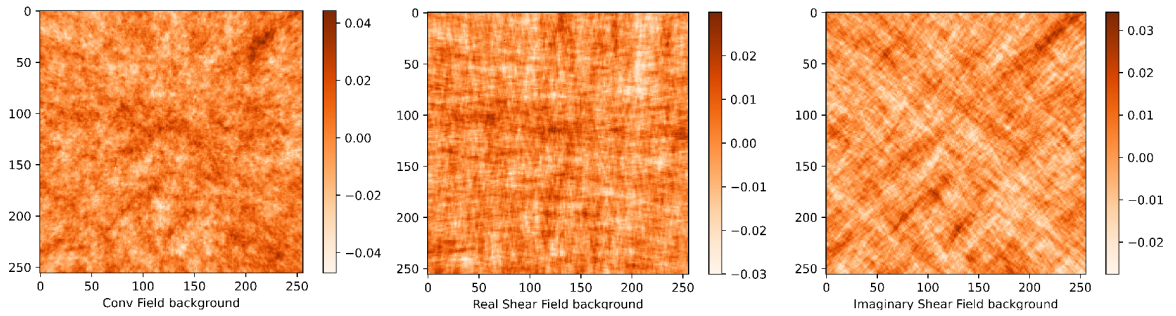


Figure 8: Convergence and shear true fields for the observer plane. Distances expressed in the axis are in Mpc. The standard deviation for convergence is  $\sigma_\kappa \sim 0.01$  and  $\sigma_{\gamma_{r/i}} \sim 0.0074$  for both shear fields. The plots suggest homogeneity as expected.

Comparing the power spectra for the convergence and shear fields will also act as a check that this integration method is being implemented successfully. As seen in Figure 12, they match within a reasonable range.

Finally, for the clustered and random sources, we used Equations 15 and 16 to add the noise originating from the unknown intrinsic ellipticities. The variance on the noise distribution is distributed equally between the real and imaginary shear parts. Because we did not have the computational power to simulate a large enough galaxy set that would counteract the large variance on the noise distribution, we decided to bring this noise down. This is effectively equivalent to increasing the number of sources.

## 4.4 REAL SPACE ANALYSIS: CORRELATION FUNCTION

This routine was used to obtain the two-point correlation function (an array) of the convergence and shear of galaxy pairs.

The algorithm considered all the possible galaxy pairs one by one and updated the element of the correlation function array corresponding to the galaxies' separation. This *bin* was then updated with a new mean, standard deviation, and the pair count was increased by one. This procedure was reiterated until all the galaxy pairs had been analysed.

This was one of the most time-consuming methods of the simulation and we tried several procedures to try and speed it up. One of the algorithms we considered involved comparing all the pairs separated at the same distance at once, by creating an array with all the possible shifts. However, if the sources were not placed on a perfect grid, this procedure could not be used. In the end, in order to keep the computational time down, we capped the galaxy pair separation to the region of interest ( $< 70\text{Mpc}$ ).

Some plots obtained using this algorithm are presented in Figure 9.

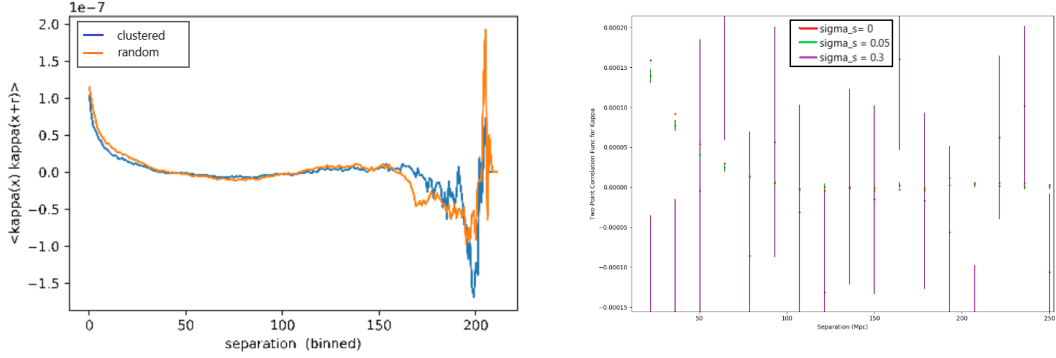


Figure 9: Correlation function graphs. Left: comparison between clustered and random sources, for which no shape noise was added.

Right (Produced by Amol): effect of the size of the shape noise on the correlation function for gridded sources. The data points appear to be *oscillating* about the true distribution ( $\sigma_s = 0$ ).

As stated in Section 3.1, we decided not to analyse our data in real space given the fact that data points in  $\xi$  are correlated. This would have involved a separate treatment which we did not consider in our report.

## 4.5 FOURIER SPACE ANALYSIS: EMPIRICAL APPROACH

The aim of this routine is to obtain the power spectrum for convergence and shear. This was our primary tool for analysing our data, since it is the method of choice in most sky-lensing surveys [15]. The algorithm involved the following steps:

### 1. Galaxy binning in real space

We created a grid of points in the source plane and we approximated the galaxies to have the position of the grid point nearest to them. This was done for clustered and random galaxies, since the gridded galaxies corresponded to the true grid point values of the convergence and shear field.

We then obtained the average of the convergence and shear signal<sup>1</sup> of all the galaxies in the same bin (grid point). As suggested by Equation 22, the standard deviation on the mean in each bin scales as  $\sim \frac{\epsilon_s}{\sqrt{N}}$ . For  $N \gg 1$ , we expect the mean convergence and shear values to approximate those of the *true* gridded sources with no error.

In the case that a bin is empty we could consider the following protocols:

- Setting the mean shear and convergence signal at that bin to 0.
- Adding one source to the bins.
- Interpolating linearly from the closest bins around.
- Using the mean shear and convergence signal value of all the bin on the source plane and using it for the empty bin value.

In our project we used the second approach, although we tried to minimise the number of empty bins by making our grid coarser.

To test our binning method, we used it on a set of gridded galaxies with no added error and we used the convergence and shear fields we had presented previously (Figure 8). If the procedure works we expect to see a similar pattern but renormalised (inspired by the Ising model [14]). The results of binning can be seen in Figure 10.

---

<sup>1</sup> Note that this is  $\epsilon_\gamma$  and  $\epsilon_\kappa$  as defined by Equations 15 and 16.



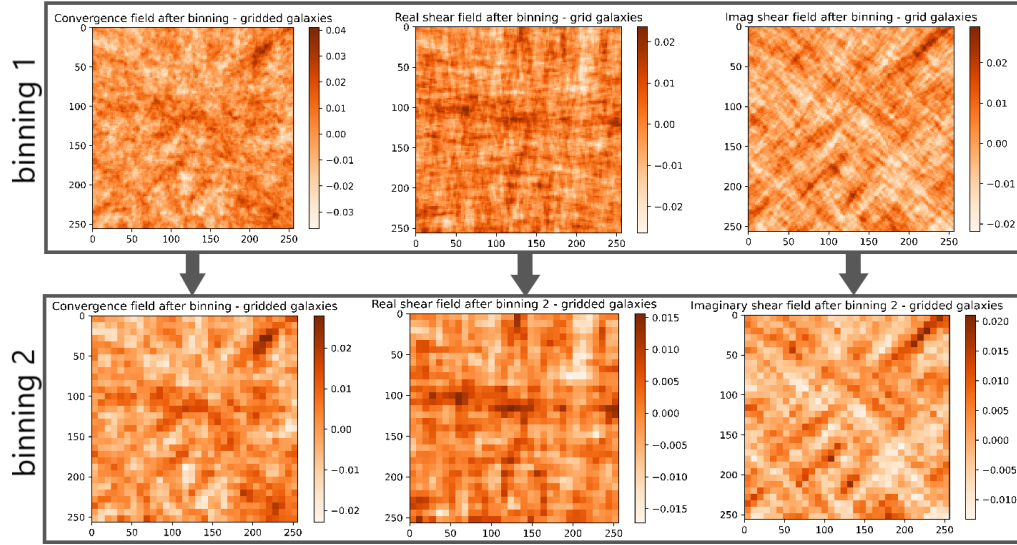


Figure 10: Convergence and shear fields of the gridded galaxies (no noise) after two iterations of binning. We see a clear similarity with Figure 8 which suggests that the binning procedure works as expected.

An alternative method, which helped verify that the clustered galaxies were being allocated to the correct bin was to look at the parameters presented in Table 1.

CLUSTERING LEVEL	EMPTY BINS	SD
random	43	2.45
low	51	2.47
medium	277	3.13
high	1365	4.89

Table 1: Table showing the number of empty bins and standard deviation (SD) on the number of galaxies per bin for different levels of source clustering. The same amount of sources have been sampled for each clustering level. The original source plane was  $256 \times 256$  Mpc<sup>2</sup> and the bins are  $2 \times 2$  Mpc<sup>2</sup>. As expected, the number of empty bins and SD increase with increasing clustering.

(Continuing with the steps of the power spectrum generation procedure)

## 2. Fourier transform of means

We simply performed a discrete Fourier transform of the 2D mean convergence and shear arrays using the *numpy* built-in routines. We obtained the Fourier amplitudes  $\tilde{\epsilon}_{\vec{k}}$ .

## 3. Obtaining the power spectrum

We firstly bin the amplitudes of the Fourier modes ( $\tilde{\epsilon}_{\vec{k}}$ ) in annuli



of phase space, where  $|\vec{k}|$  ranges from  $|\vec{k}| \rightarrow |\vec{k}| + \delta k$ . The shape of this annuli can be seen in Figure 11.

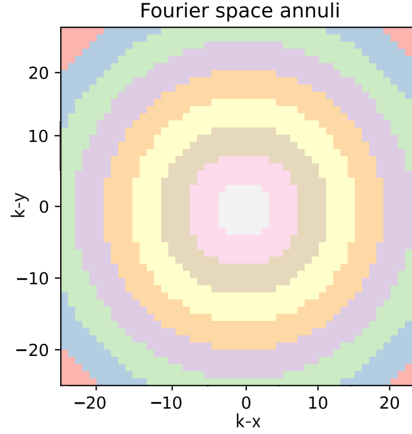


Figure 11: Shape of the Fourier annuli. In this example sketch we have 9 bins for a Fourier space of  $50 \times 50 \text{ Mpc}^{-2}$ . We notice that the annuli corresponding to small and large  $|\vec{k}|$  contain fewer amplitudes in them, so we expect larger errors on the mean.

We calculated the power spectrum value for each bin with the following expression:

$$P_{k+\delta k} = \frac{1}{g_k} \sum_n |\tilde{\epsilon}_{\vec{k}}(\vec{k}_n)|^2 \quad (43)$$

where  $g_k$  is the number of amplitudes whose  $|\vec{k}|$  lies within the annulus. The standard error of the mean (SEM) is worked out by collecting all the  $|\tilde{\epsilon}_{\vec{k}}(\vec{k}_n)|^2$  values of the annulus and calculating its standard deviation using built-in Python statistics packages (*scipy.stats*).

An example plot for the shape of the convergence and shear power spectra is depicted in Figure 12.

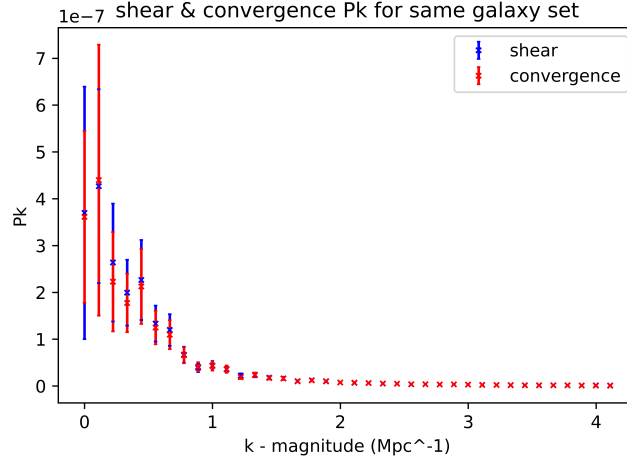


Figure 12: Convergence and shear power spectrum for the same galaxy set. As expected, they seem to match reasonably well. Deviations between them could be due to an edge effect, the expectation that they are equal only holds in the infinite field limit. Alternatively, this could also be a result of our limited computational power.

In order to be able to compare the Power spectra for different sources, we needed to obtain the variance of the Fourier shape noise,  $\tilde{\sigma}_k^2$  (see Section 3.4). This variance produces the *shot noise* effect which, *shifts* all be values in the power spectrum by the same amount. This phenomenon can be observed in Figure 13 for different intrinsic shape distributions.

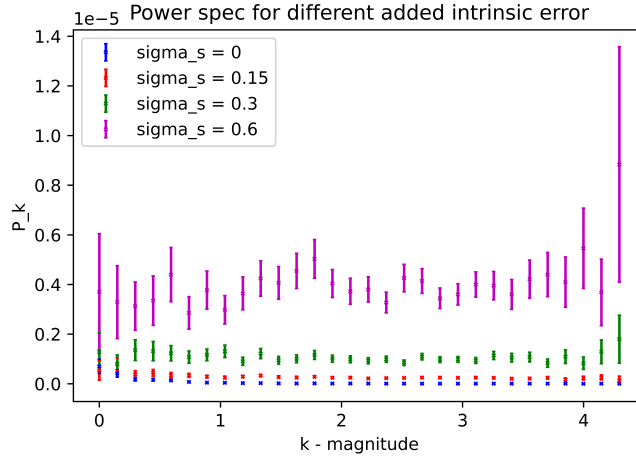


Figure 13: The effect of the size of the intrinsic shape distribution,  $\sigma_s$  on power spectrum's shape noise variance,  $\tilde{\sigma}_k^2$ . As expected from theory,  $\tilde{\sigma}_k^2 \propto \sigma_s^2$ , which proved the validity of our routine.  $\sigma_s = 0$  corresponds to the *true* gridded sources.

The methods we used to calculate the shot noise were:

- Analytically:  $\tilde{\sigma}_k^2 = \sum_b \sigma_b^2$ , with  $\sigma_b^2 = \frac{\sigma_s^2}{g}$  (Equation 24). This requires all bins to have at least one source.

- Empirically: we run the same galaxy set several times ( $\sim 30$ ), creating different random noise each time (from the same distribution,  $\sigma_s$ ). In Fourier space we then calculate the variance on the shear and convergence for each  $\vec{k}$ .
- Empirically: we run the same galaxy set twice, once with added noise and once without. We then obtained the average separation between their power spectra (we use the tail region where the power spectrum has stabilised), this is  $\tilde{\sigma}_{\vec{k}}$ .

These three procedures were all in agreement with each other. However, we primarily used the second empirical method since in some of our tests not all bins were occupied and we did not have a large enough source sample.

#### 4.6 FOURIER SPACE ANALYSIS: ANALYTICAL APPROACH

In this section we used our derivations from Section 3.4 in the THEORY Chapter. This procedure was mainly used to support our empirical power spectrum findings.

We decided to consider the following alternative approach because we realised that the amplitudes of the Fourier modes corresponding to the same  $|\vec{k}|$  had a significant variation between them (see Figure 14), which increased the error on the mean power spectrum value for each annuli.

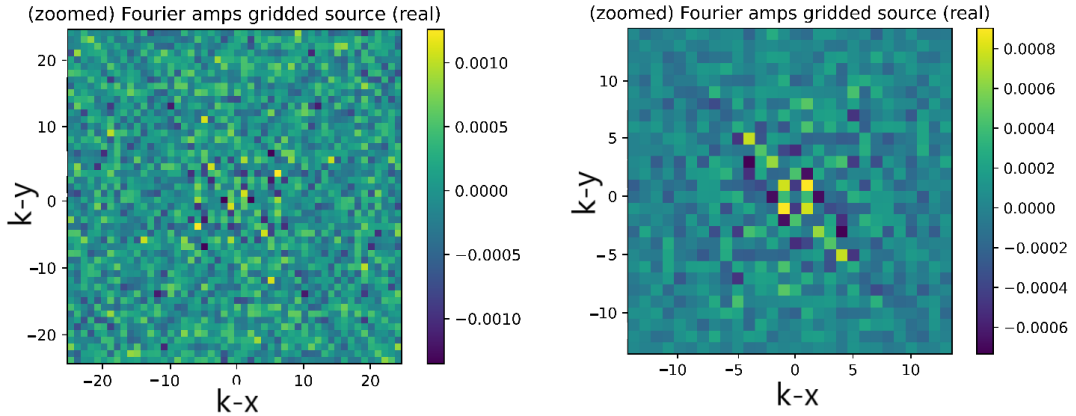


Figure 14: Real side of the power spectrum for a set of gridded sources (zoomed at different scales). It is clear that different  $\vec{k}$  that fall within the same  $|\vec{k}| \rightarrow |\vec{k}|+$  vary significantly in value.

This, together with the small size of the k-space annuli for very large and very small  $\vec{k}$  (Figure 11) was our motivation to test this method. This analytical investigation was, however, limited due to our lack of knowledge of the true value for  $\tilde{\sigma}_{\vec{\mu}}$ , for which we only had an estimate <sup>2</sup>.

<sup>2</sup> We could have tried using a large enough sky simulation and take the estimate of the value of  $\hat{\sigma}_{\vec{\mu}}$  to be the true value  $\hat{\sigma}_{\vec{\mu}} = \tilde{\sigma}_{\vec{\mu}}$ , and use this in a smaller scale simulation.

In order to be able to use our analytical tools, we filled any empty bins with one galaxy <sup>3</sup>. The parameters we analytically obtained were:

- the variance of the noise in Fourier space,  $\tilde{\sigma}_{\vec{k}}^2$ , using Equation 24.
- the estimation for the total variance,  $\hat{\sigma}_T^2$ , using Equation 29, which is equivalent to  $P_{k+\delta k}$  in Equation 43.

With these parameters available we decided to take two different approaches to investigating the Fourier field analytically:

#### 1. Comparison of the Fourier modes amplitudes

With this approach we aim to compare the amplitudes for each  $\vec{k}$  individually.

First, we Fourier-transform the binned convergence and shear fields for the clustered and random sources, as per the previous section, and we also calculate  $\tilde{\sigma}_{\vec{k}}^2$ . In order to compare the Fourier amplitudes of the clustered and random sources to the *true* gridded sources, we calculate the chi squared value (see Section 3.3):

$$\chi^2 = \sum_{\vec{k}} \frac{|\tilde{\epsilon}_{\vec{k}} - E[\tilde{\epsilon}_{\vec{k}}]|^2}{\text{Var}[\tilde{\epsilon}_{\vec{k}}]} = \sum_{\vec{k}} \frac{|\tilde{\epsilon}_{\vec{k}} - \tilde{\mu}_{\vec{k}}|^2}{\tilde{\sigma}_{\vec{k}}^2} \quad (44)$$

This procedures are easily computed using the *numpy* methods for arrays.

#### 2. Comparison of the power spectrum signal-to-noise ratio

Using the estimator  $\hat{\sigma}_T^2$ , which we have already obtained in the empirical treatment, we will find the ratio of signal to noise, as given in Equation 34. We expect the gridded sources with no added intrinsic shape error to have provide the smallest value, since  $\tilde{\sigma}_{\vec{k}}^2 = 0$  in their case.

Overall the computational procedure was similar to that of the Fourier empirical analysis (Section 4.5).

<sup>3</sup> To see how the protocol for fillig empty bins affects the shape noise and the noise-to-signal ratio, refer to Amol Bhuptani's report.

## RESULTS

For the final results that will be presented below we generated a gravitational potential of dimensions:  $256 \times 256 \times 1280 \text{ Mpc}^3$ . The true convergence and shear fields are those in Figure 8.

We created 6 different galaxy sets with approximately the same number of sources,  $\sim 100,000$ , providing a source density of  $\sim 1.5 \frac{\text{sources}}{\text{Mpc}^2}$ . The types of sets created were: 4 clustered, 1 random and 2 gridded (with and without noise). The intrinsic shape error of the galaxies was drawn from the same  $\sigma_s$ , which we set to be 0.15. The characteristics of these galaxies are described in Table 2. Figure 15 shows heatmaps for the number of galaxies per bin of the random, C3 and C4 galaxy sets.

CLUSTERING LEVEL	$\sigma_\delta$	SD(BIN)
random	0.0	4.93
C1	0.05	5.00
C2	0.12	5.11
C3	0.50	7.86
C4	1.20	13.95

Table 2: Table showing the standard deviation of the overdensity field ( $\sigma_\delta$ ) that generated the galaxies, and standard deviation (SD) on the mean number of galaxies per bin after binning. C1  $\rightarrow$  C4 are the galaxy sets codes indicating increasing level of clustering. The original source plane was  $256 \times 256 \text{ Mpc}^2$  and the bins are  $4 \times 4 \text{ Mpc}^2$ . Given that the number of galaxies was  $\sim 100,000$ , there was a mean of 24.5 galaxies per bin.

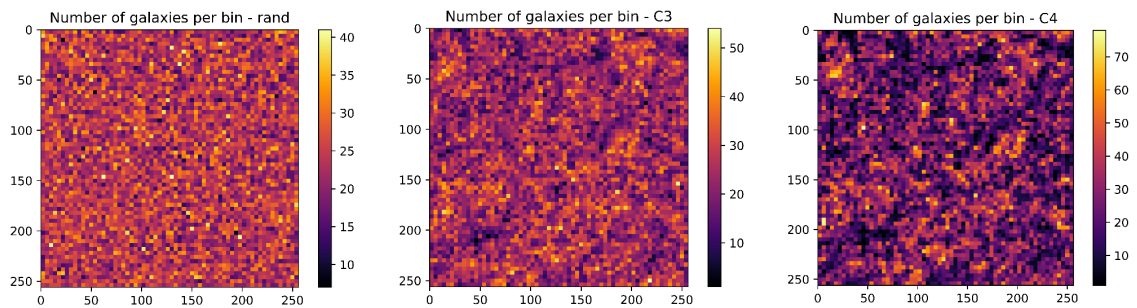


Figure 15: Number of galaxies per bin ( $4 \times 4 \text{ Mpc}^2$ ) for (left to right) random, C3 and C4. There is an apparent increase in clustering from left to right as expected.

The only set to produce empty bins was C<sub>4</sub> (74 empty bins), to which we added one source. The binning in  $k$ -space was done using 40 annuli. We chose to have a small number of annuli to increase the modes density and get lower errors on the mean for  $\hat{\sigma}_{\mu}^2$ . The number of Fourier modes per bin is show in Figure 16.

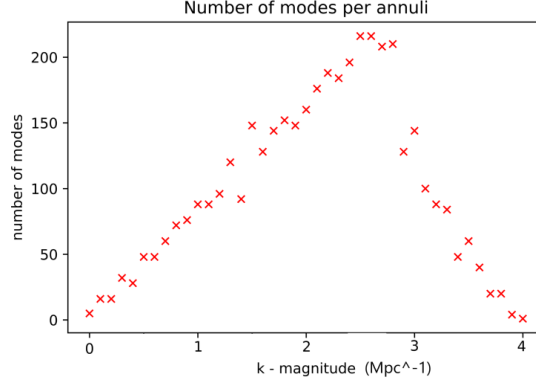


Figure 16: Number of Fourier modes per annulus. This plot is expected given the shape of the bins as seen in Figure 11. The mean is 100.

### 5.1 EMPIRICAL POWER SPECTRA RESULTS

This section presents the most relevant power spectra results we obtained empirically. Figure 17 depicts the spectra comparison for galaxy sets C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub> and C<sub>4</sub>.

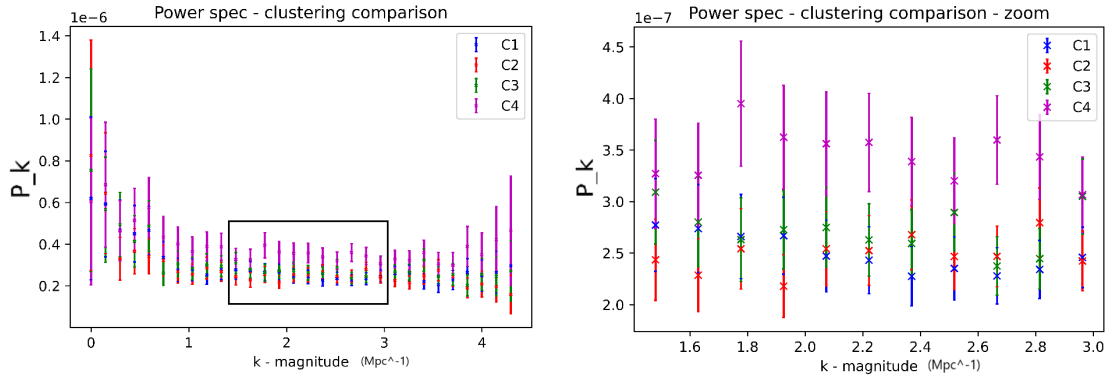


Figure 17: Left: power spectra for different levels of clustering. Right: zoomed on smaller section.

These plots show a larger shape noise,  $\tilde{\sigma}_{\vec{k}}$ , and larger error bars,  $\text{Var}[\hat{\sigma}_{\mu}^2]$ , for increasing clustering intensity.

It was also interesting to observe the average standard deviation on the mean of the power spectrum signal, shown in Table 3.

CLUSTERING LEVEL	STD( $P(\bar{k})$ ) ( $10^{-8}$ )
gridded (with noise)	2.227
random	8.298
C1	8.5858
C2	8.6906
C3	9.2456
C4	11.5823

Table 3: Mean standard deviation (error) on the mean of the power spectrum signal in the annuli. Increased clustering is associated with larger error bars.

Our estimations on the shot noise value ( $\tilde{\sigma}_k^2$ ) for the different galaxy sets are presented in Table 4.

CLUSTERING LEVEL	$\tilde{\sigma}_k^2$ ( $10^{-7}$ ) FROM $\kappa$	$\tilde{\sigma}_k^2$ ( $10^{-7}$ ) FROM $\gamma$
gridded (with noise)	2.22123	2.22041
random	2.2452	2.2446
C1	2.3784	2.3791
C2	2.4586	2.4573
C3	2.7035	2.7021
C4	3.4312	3.4291

Table 4: Empirical estimates we obtained for the variance of the shape noise. These were worked out by generating the same galaxy sets twice, one with added intrinsic shape noise and one without, and obtaining the average difference. The variance obtained from the convergence and shear power spectrum are in agreement since they are essentially equivalent fields, as seen from Figure 12.

We visually and statistically compared the power spectrum of the gridded galaxies against all the other sets, adding the shot noise values from Table 4 to the *true* gridded galaxies. This was essentially equivalent to *removing* the shape noise from the clustered galaxies. Table 5 shows the chi squared values per degree of freedom obtained from comparing these *true* and observed power spectra. Figure 18 presents comparative plots of the spectra of the clustered and gridded galaxies.

CLUSTERING LEVEL	$\chi^2/N_{\text{dof}}$ FROM $\kappa$	$\chi^2/N_{\text{dof}}$ FROM $\gamma$
random	1.10423	0.92294
C1	1.09172	0.90156
C2	0.87434	0.96621
C3	0.93811	1.10421
C4	0.91895	0.89983

Table 5: Chi squared values per degree of freedom obtained using the power spectrum of the gridded galaxies (with added shot noise) as the theoretical expectation, and the power spectrum of the clustered galaxies as the observed data. They all lie within the expected  $\chi^2$  value per degree of freedom ( $N_{\text{dof}} = 29 \rightarrow \chi^2 = 1 \pm 0.2626$ ). This shows that clustering does not cause a systematic deviation from the true value.



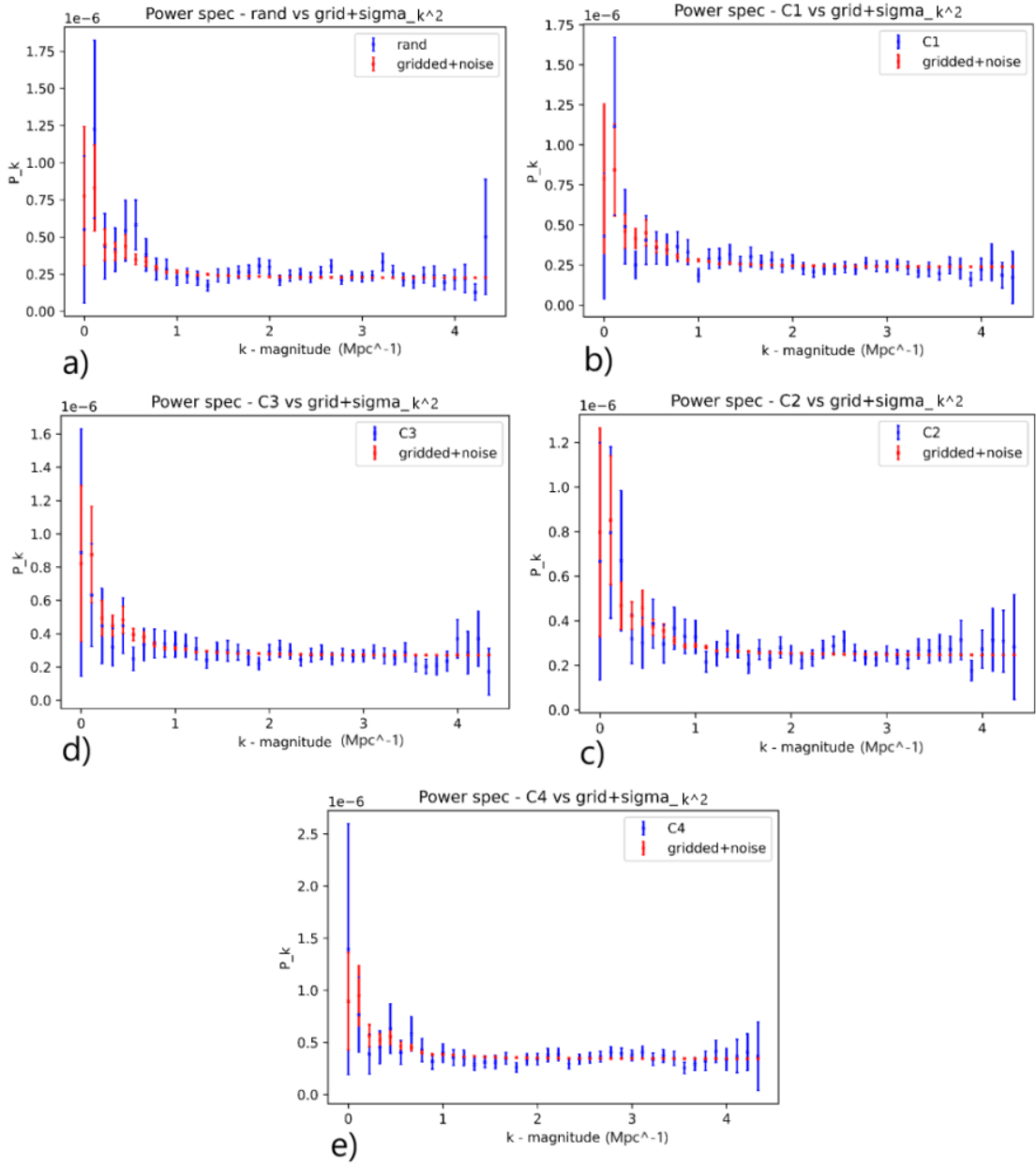


Figure 18: Plots of the power spectrum for convergence of all the galaxy sets versus the true field (gridded sources). The gridded source have been added the shape noise value of the galaxy set they are being compared against. These plots suggest that our estimates from the shape noise are accurate.

## 5.2 FOURIER AMPLITUDE COMPARISON RESULTS

The chi squared values calculated using Equation 44 for all our galaxy sets against the gridded *true* field are shown in Table 6.

CLUSTERING LEVEL	$\chi^2/N_{\text{dof}}$ FROM $\kappa$	$\chi^2/N_{\text{dof}}$ FROM $\gamma$
random	1.01244	1.01829
C1	1.00920	0.98940
C2	0.98473	1.00853
C3	0.98067	1.01648
C4	0.98896	0.98887

Table 6: Chi squared values obtained from comparing the Fourier amplitudes for all the  $\vec{k}$  directly against those of the gridded sources. They all lie within the expected  $\chi^2$  value per degree of freedom ( $N_{\text{dof}} = 4096 \rightarrow \chi^2 = 1 \pm 0.0221$ ). Once again this demonstrates that clustering does not cause a systematic deviation from the true value.

### 5.3 SIGNAL-TO-NOISE RATIO

We obtained the signal-to-noise values as presented in Section 3.4 with Equation 34. For this observation I will include the graph generated by Amol, whose simulation had significantly smaller error bars due to the larger number of sources he simulated. This graph can be found below in Figure 19.

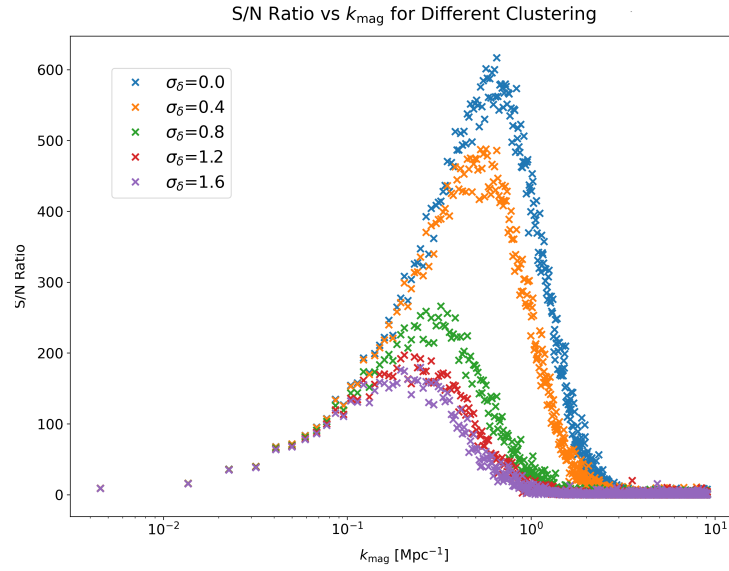


Figure 19: (Generated by Amol Bhuptani) Signal-to-noise (SN) ratio for the power spectrum of galaxies with different clustering.  $\sigma_\delta$  indicates the standard deviation of the matter overdensity field. It is clear that more intense clustering leads to a lower SN ratio. This figure is only valid when the empty bin filling protocol was merely adding one source per bin, like we did for this simulation. It is interesting to notice the turning point in the distributions, most likely caused by the turning point in the number of modes per annuli (Figure 16), which leads to a worse estimate on the mean.

## CONCLUSIONS

---

The effect of source clustering on survey measurements is still an unexplored field in the arena of weak gravitational lensing. Despite the short time frame, this first attempt to numerically quantify the significance of this previously neglected phenomenon has been successful in assessing the basic effects of source clustering on Fourier space parameters (power spectra).

From a modelling perspective, within our computational limitations, we managed to create a robust environment for weak lensing simulations, capturing the main features of source clustering with reduced running times. The tests carried out alongside the writing of the code indicate that the methods have been implemented successfully.

Our empirical and analytical comparative results, namely the  $\chi^2$  values presented in Tables 5 and 6, suggest that clustering does not cause a systematic deviation from the true value ( $\chi^2 \sim N_{\text{dof}}$ ). It does, however, produce larger variance on the shape noise,  $\tilde{\sigma}_k^2$ , which increases with stronger clustering, as seen in Table 4. We expected this from theory, since the real space binned configuration to minimise this noise is that corresponding to equally spread (i.e. gridded) galaxies. This proportionally larger noise is also evident when comparing the signal-to-noise ratio for different levels of clustering. Increased uncertainties are also observed in the error on the mean of the power spectrum signal of each annuli (Table 1). Our results suggest that these larger errors are especially significant for  $\sigma_\delta > 0.4$ .

More work needs to be done in order to quantify the effects of clustering in the estimation of cosmological parameters, which is the ultimate aim of the upcoming high precision surveys. This would have been our next area of investigation had we had more time available. Moreover, in our simulation, we would have liked to implement redshift dependence on the deflection of light.

Our simulation was subject to strong assumptions and approximations, which should be revisited for a more realistic analysis. The effects of some of these neglected phenomena are more significant than the implications of source clustering alone. We suggest future studies consider the combined effect of intrinsic alignment (IA) and clustering, which could potentially be more substantial than source clustering independently. Biases produced by IA alone have been explored more in depth by previous studies [16,17,18] which have shown that, if ignored, it can

alter cosmological inferences by tens of per cent [1]. We expect source clustering to amplify the effects of IA.

## BIBLIOGRAPHY

---

- [1] Martin Kilbinger, *Cosmology with cosmic shear observations: a review*. CEA Saclay, 2015. DOI: 10.1088/0034-4885/78/8/086901.
- [2] *Overview of the SuperNova/Acceleration Probe (SNAP)*, 2002. arXiv:astro-ph/0209550
- [3] Bartelmann and Schneider, *Weak Gravitational Lensing*. Max-Planck-Institut fur Astrophysik, 2008.
- [4] Munshi, Valageas, van Waerbeke and Heavens, *Cosmology with Weak Lensing Surveys*, 2018.
- [5] Hyper Supreme-Cam Subaru Strategic Programme, *Weak Lensing Cosmology*, 2016-2019.
- [6] Huterer, Takada, Bernstein and Jain, *Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration* , 2006. DOI: 10.1111/j.1365-2966.2005.09782.x
- [7] Roberto Trotta, lecture notes for the course: *Physics of the Universe: Cosmology and Astroparticle Physics*, Imperial College London, 2020.
- [8] Max Tegmark et al. *Cosmological parameters from SDSS and WMAP*, 2004. DOI: 10.1103/PhysRevD.69.103501.
- [9] Michael Norman, *Simulating galaxy clusters*, Department of Physics, UC San Diego.
- [10] Carlo Contaldi, lecture notes on the course: *Fourier Analysis*, Imperial College London, 2019.
- [11] Jones, Martinez, Saar and Trimble, *Scaling laws in the distribution of galaxies*, 2004.
- [12] Mark Richards, lecture notes on the course *Statistics of measurement*, Imperial College London, 2019.
- [13] Scott Dodelson, *Gravitational Lensing*, Cambridge University Press, 2017. DOI: 10.1017/9781316424254.
- [14] Christensen and Moloney, *Complexity and Criticality*, Imperial College Press, 2005. DOI: 10.1142/9781860949319/0002.
- [15] Hu and Tegmark, *Weak Lensing: Prospects for Measuring Cosmological Parameters*, Institute for Advanced Study, Princeton, published on The Astrophysical Journal, 1999.

- [16] Wei et al, *Exploring Galaxy Intrinsic Alignment and Cosmic Shear Correlations*, The American Astronomical Society, 2018. DOI: 10.3847/1538-4357/aaa4od.
- [17] Heavens, Refregier and Heymans, *Intrinsic correlation of galaxy shapes: implications for weak lensing measurements*, 2000.
- [18] Troxel and Ishak, *The Intrinsic Alignment of Galaxies and its Impact on Weak Gravitational Lensing in an Era of Precision Cosmology*, Physics Reports, 2014.

